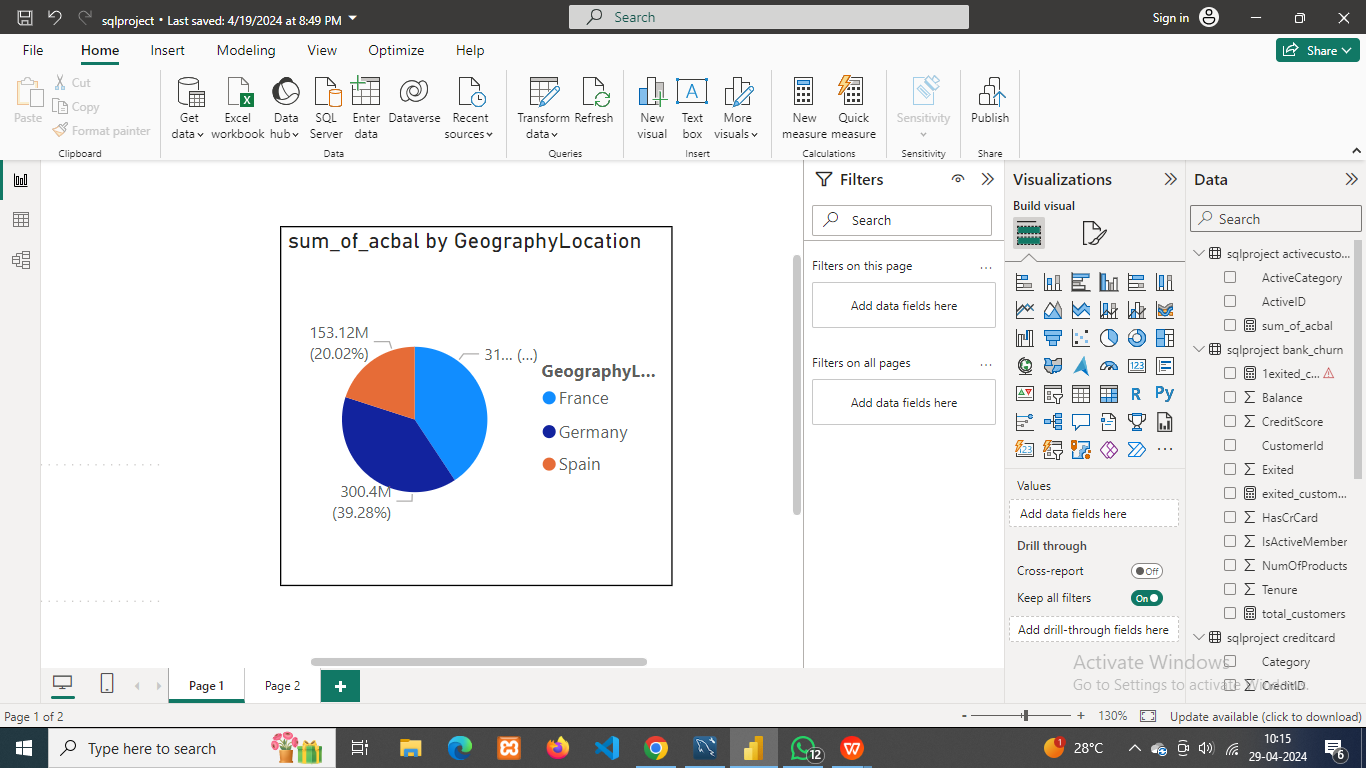
BY MALLESHA H

Objective Questions

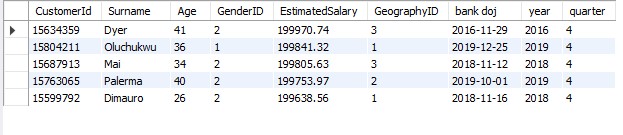
1.What is the distribution of account balances across different regions?

****

The distribution of account balances in different regions as follows

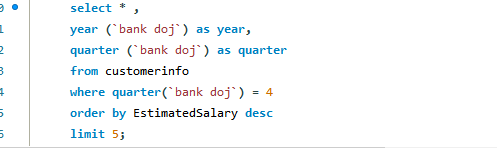
* France 311.33 millions
* Germany 300.4 millions
* Spain 153.12 millions

1. Identify the top 5 customers with the highest Estimated Salary in the last quarter of the year. (SQL)

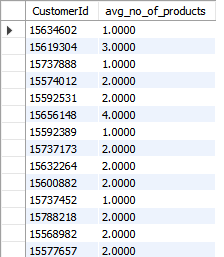


Displayed in the table are the top 5 customers with the highest estimated salaries, all belonging to the final quarter of the year.

SQL query:

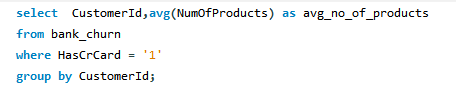


1. Calculate the average number of products used by customers who have a credit card. (SQL)



The above table displays average number of products used by each customers who have a credit card.

SQL query



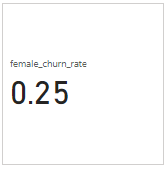
1. Determine the churn rate by gender for the most recent year in the data set.

The most recent year in the given data set is 2019 and the churn rate of male is 0.16



Formula used is:male\_chrn\_rate = CALCULATE(([male\_cus\_exited]/[Total\_male\_customers]),'sqlproject customerinfo'[year])

Churn rate of female is 0.25



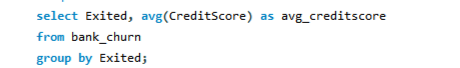
Female churn rate =female\_chrn\_rate = CALCULATE(([female\_cus\_exited]/[Total\_male\_customers]),'sqlproject customerinfo'[year])

1. Compare the average credit score of customers who have exited and those who remain. (SQL)



Average credit sore for customers who remain is greater than the customer who exited

SQL query

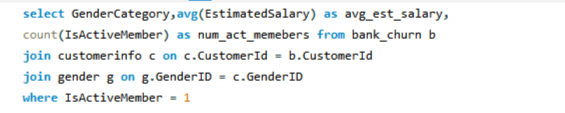


1. Which gender has a higher average estimated salary, and how does it relate to the number of active accounts? (SQL)

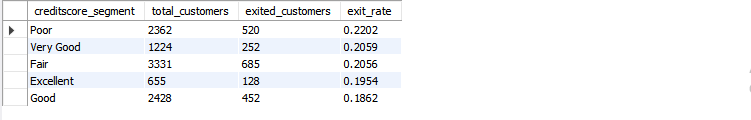
it's evident that the Female category boasts a higher average estimated salary compared to the Male category. Interestingly, this disparity doesn't correlate with the number of active accounts. Despite having fewer active accounts, the Female category maintains a higher average estimated salary, contrasting with the Male category's larger active account count.



SQL query

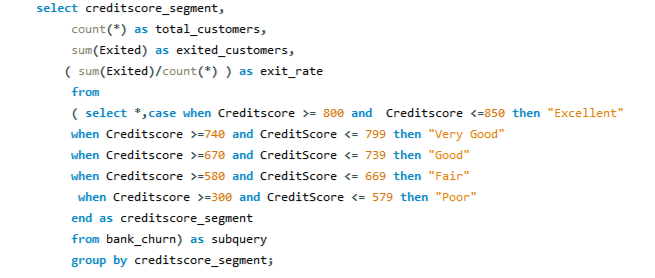


1. Segment the customers based on their credit score and identify the segment with the highest exit rate. (SQL)



Poor segment has the highest exit rate 0.2202

SQL query

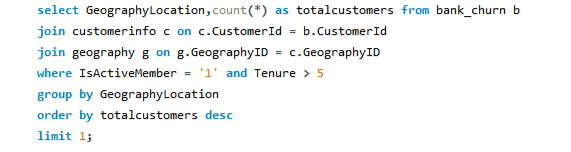


1. Find out which geographic region has the highest number of active customers with a tenure greater than 5 years. (SQL)

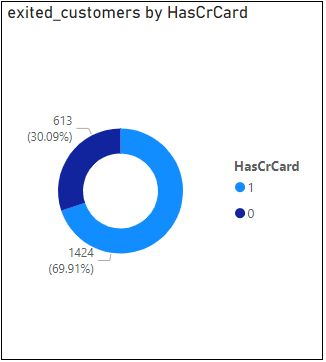
France has the highest number of active members



SQL query

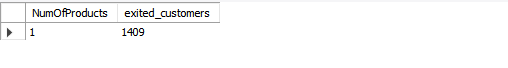


1. What is the impact of having a credit card on customer churn, based on the available data?



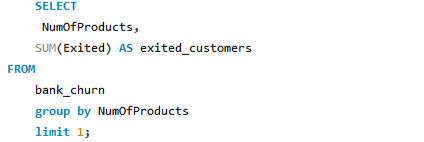
The chart above suggests a higher likelihood of customer churn among those with credit cards, implying potential dissatisfaction with our credit card services. To address this, it's essential to explore and enhance our credit card offerings to improve customer satisfaction and retention.

1. For customers who have exited, what is the most common number of products they have used?

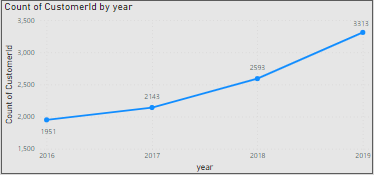


1 is the number of products which is the most common number of products used by the exited customer.

SQL query

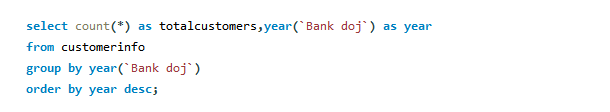


1. Examine the trend of customers joining over time and identify any seasonal patterns (yearly or monthly). Prepare the data through SQL and then visualize it



The trend of customer growth is noticeable year over year. Specifically, in 2019, we observed the highest influx of customers, indicating significant expansion or heightened interest

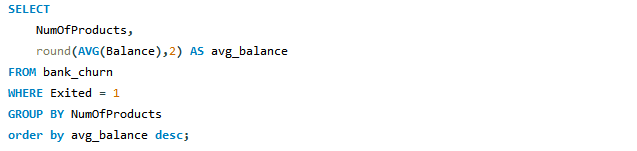
SQL query



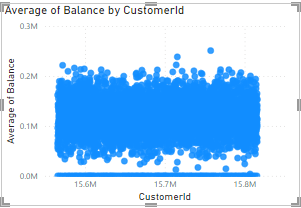
1. Analyze the relationship between the number of products and the account balance for customers who have exited.



SQL query



1. Identify any potential outliers in terms of balance among customers who have remained with the bank.



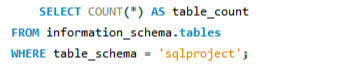
The visual reveals a disparity in account balances among customers who have remained with the bank. Some accounts exhibit significantly high balances, while others show minimal balances, nearing zero. These outliers underscore the wide range of balances within the customer base, highlighting instances of both substantial wealth and more modest financial situations among our retained customers.

1. How many different tables are given in the data set, out of these tables which table only consists of categorical variables?

There are seven different tables in the given data set.



SQL query

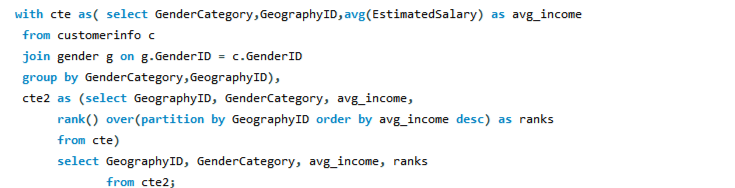


Tables which consists categorical variables are : activecustomer, creditcard, exitcustomer, gender and geography.

1. Using SQL, write a query to find out the gender-wise average income of males and females in each geography id. Also, rank the gender according to the average value. (SQL)



SQL query

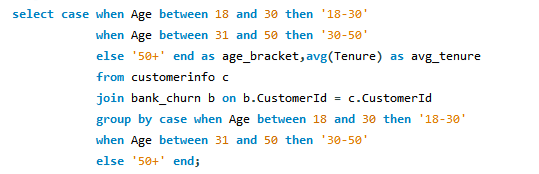


Based on the provided table, we can derive the average income for each category across different geographic locations. Moreover, within each geographic location, the gender categories are organized based on income, revealing the income rankings within their respective categories. This offers insights into income distribution and gender-based income disparities across different regions.

1. Using SQL, write a query to find out the average tenure of the people who have exited in each age bracket (18-30, 30-50, 50+).



SQL query



1. Is there any direct correlation between salary and the balance of the customers? And is it different for people who have exited or not?



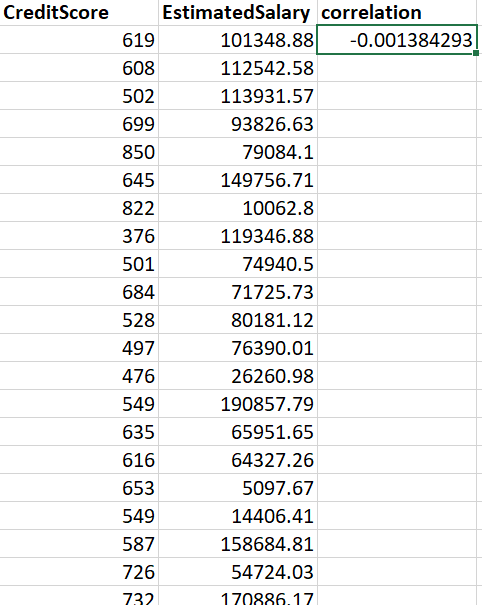
The analysis indicates a weak correlation between customer salary and balance, with a correlation coefficient close to zero, suggesting a lack of direct relationship. Specifically among exited customers, a negative correlation was found between salary and balance, indicating an inverse relationship: higher salaries correspond to lower balances, and vice versa.

1. Is there any correlation between the salary and the Credit score of customers?

There is a correlation between the salary and the credit score of customers

**Correlation=-0.001384293**

Excel formula used= =CORREL(K2:K10001,L2:L10001)

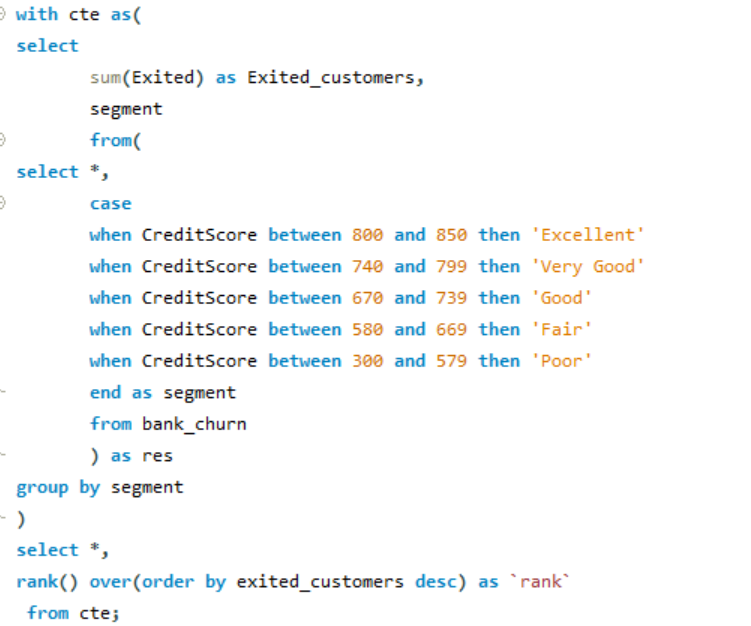


1. Rank each bucket of credit score as per the number of customers who have churned the bank.



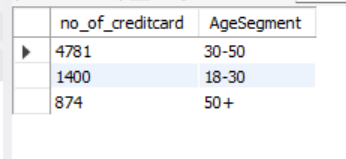
The table ranks each credit score bucket based on the number of customers who have churned from the bank.

SQL query

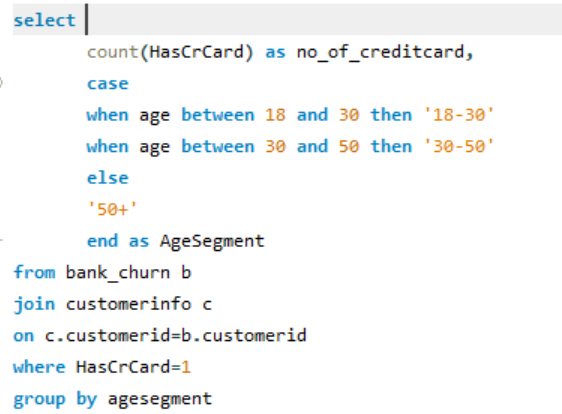


1. According to the age buckets find the number of customers who have a credit card. Also retrieve those buckets who have lesser than average number of credit cards per bucket.

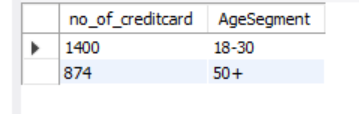
According to the age buckets number of customers having credit card are



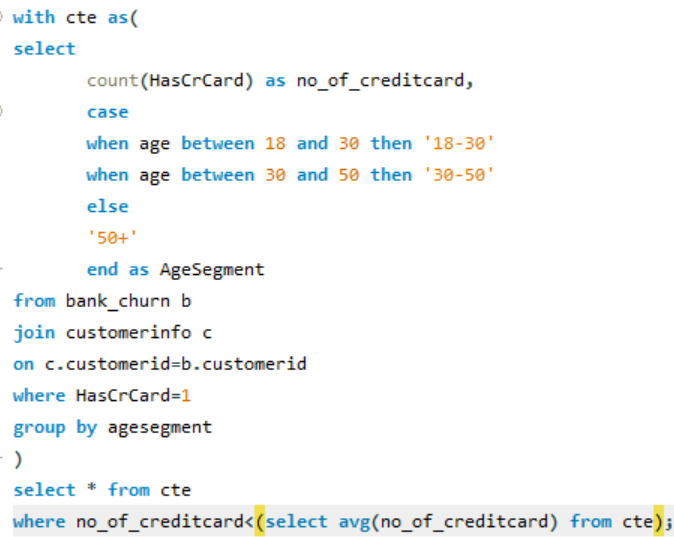
SQL query



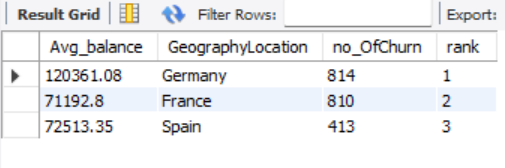
Age buckets which have lesser than average number of credit cards per bucket are



SQL query

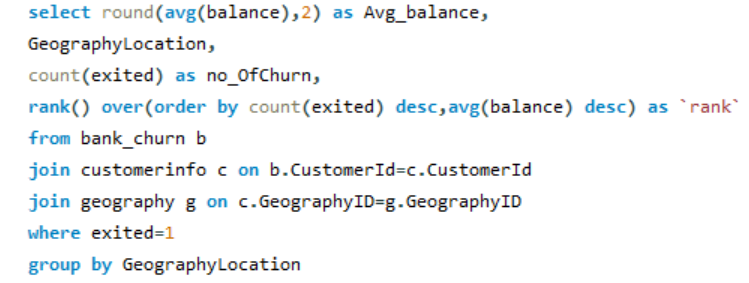


1. Rank the Locations as per the number of people who have churned the bank and average balance of the customers.



From the above table, we can identify the locations with the highest and lowest ranks based on the number of people churned and the average balance of customers.

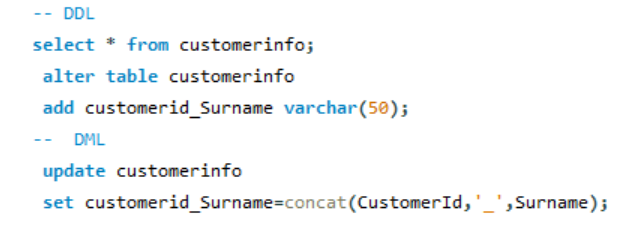
SQL query



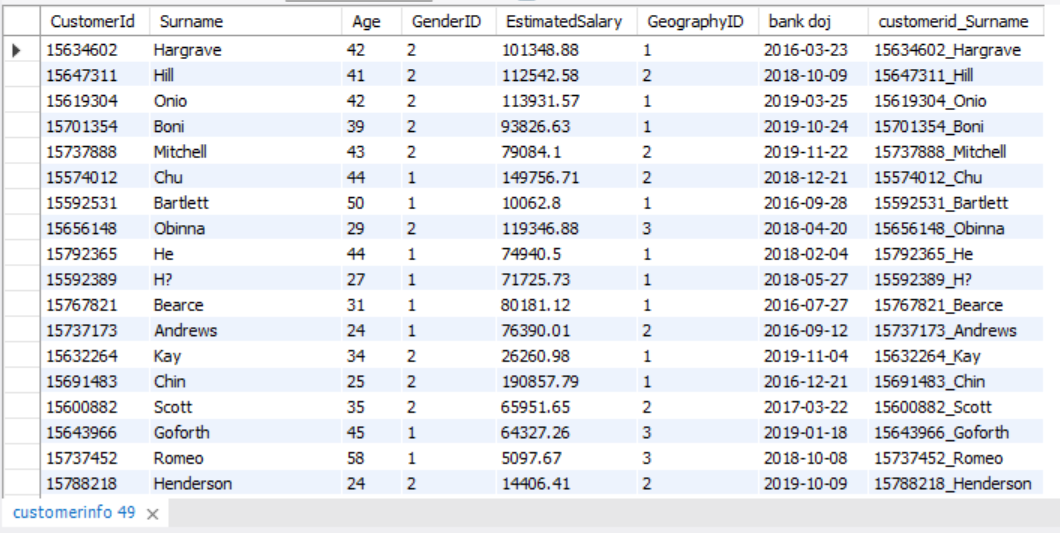
1. As we can see that the “CustomerInfo” table has the CustomerID and Surname, now if we have to join it with a table where the primary key is also a combination of CustomerID and Surname, come up with a column where the format is “CustomerID\_Surname”.

To achieve this task, I started by modifying the table structure using Data Definition Language (DDL) to add a new column called CustomerID\_Surname. Following that, I utilized Data Manipulation Language (DML) to update the records, ensuring that the new column was populated with the desired format "CustomerID\_Surname", which merged the CustomerID and Surname values.

SQL query

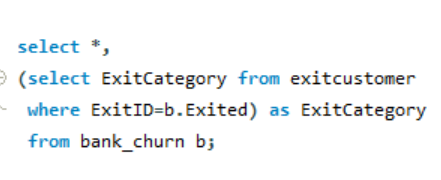


Output:



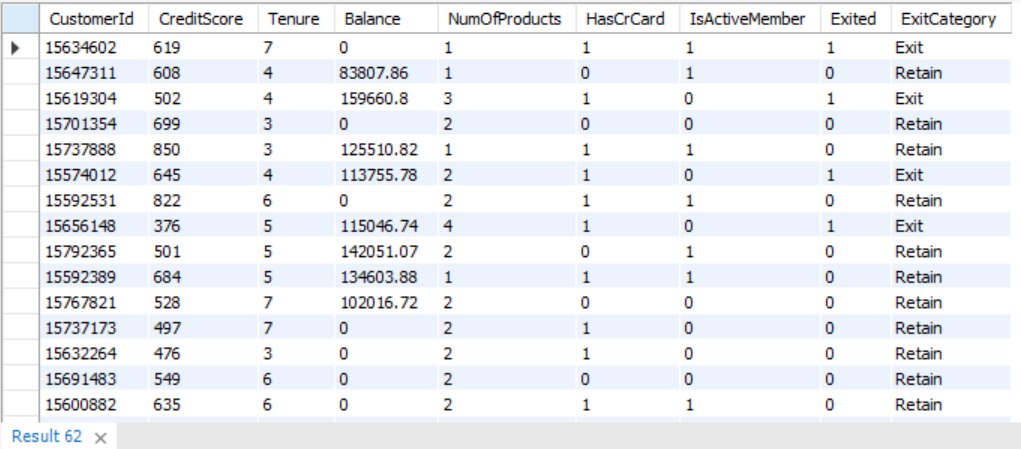
1. Without using “Join”, can we get the “ExitCategory” from ExitCustomers table to Bank\_Churn table? If yes do this using SQL.

Without using join, we can retrieve the "ExitCategory" from the ExitCustomers table and incorporate it into the Bank\_Churn table by employing a subquery.

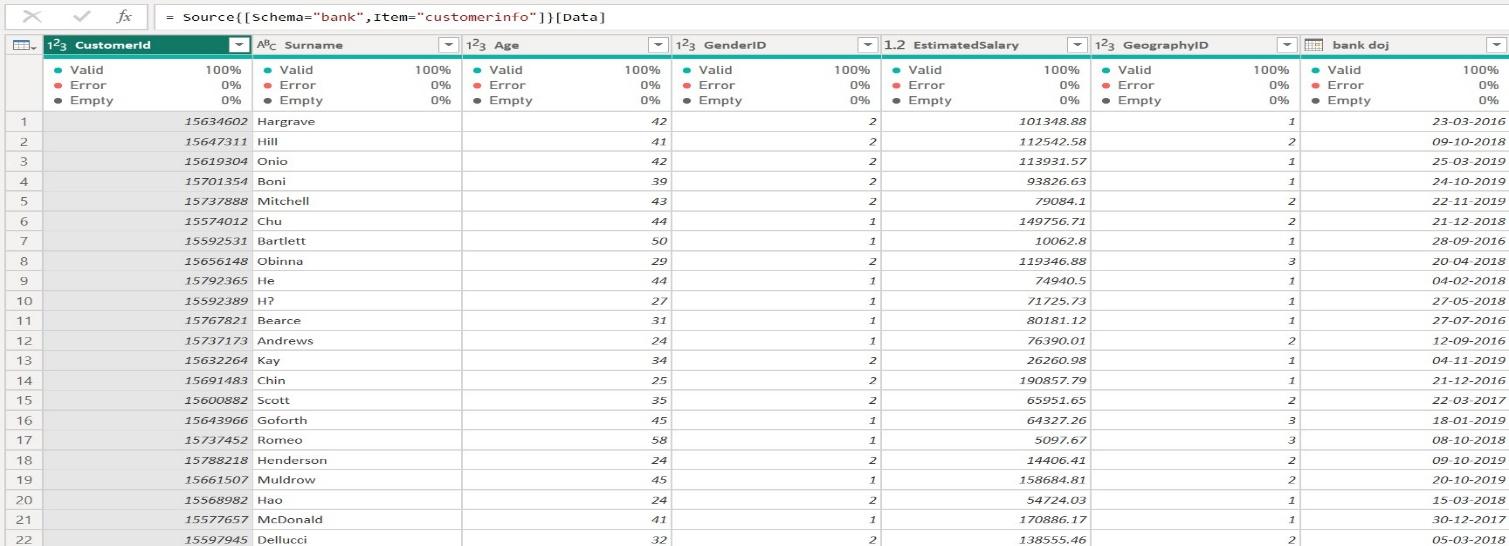


The query selects all columns from the **bank\_churn** table while adding a new column named **ExitCategory**. A subquery is utilized within the main query to retrieve the **ExitCategory** from the **exitcustomer** table, where the **ExitID** matches the **Exited** field in the main query's result set. This approach seamlessly incorporates the **ExitCategory** information from the **exitcustomer** table into the **bank\_churn** table without employing a join operation.

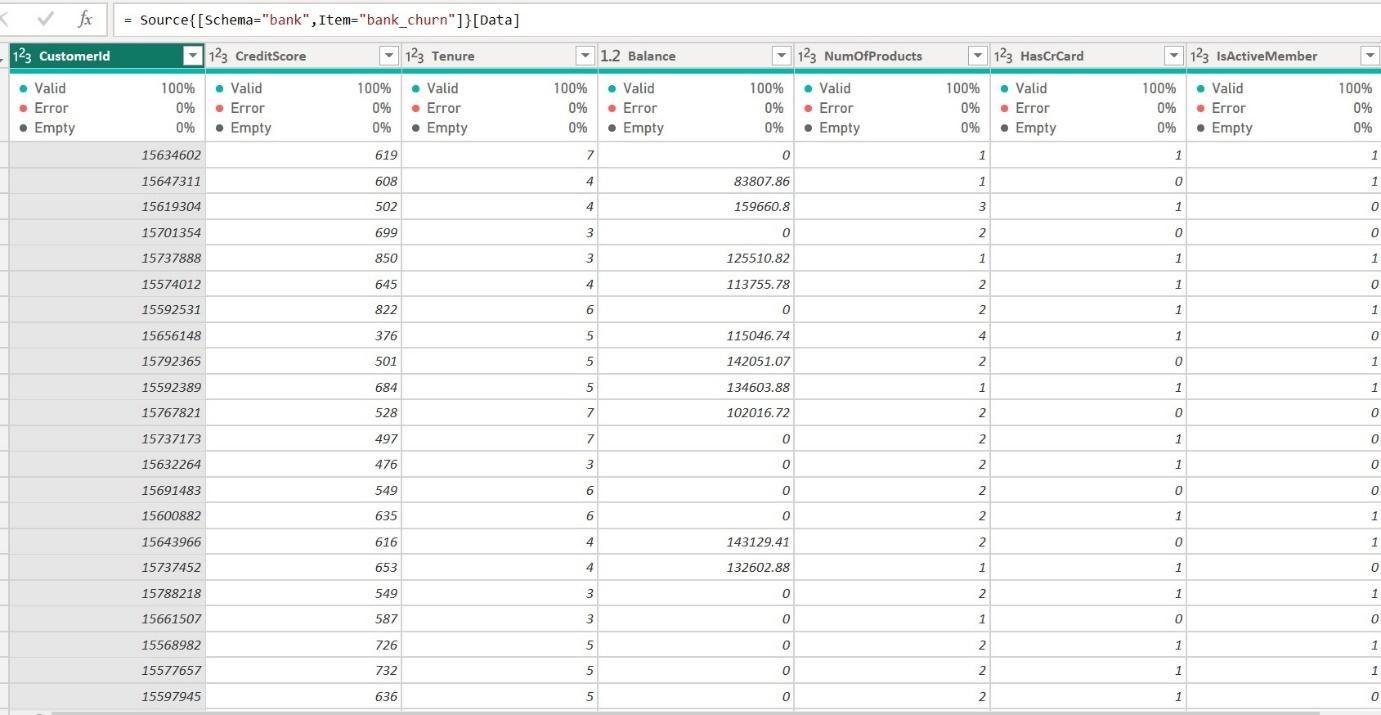
Output

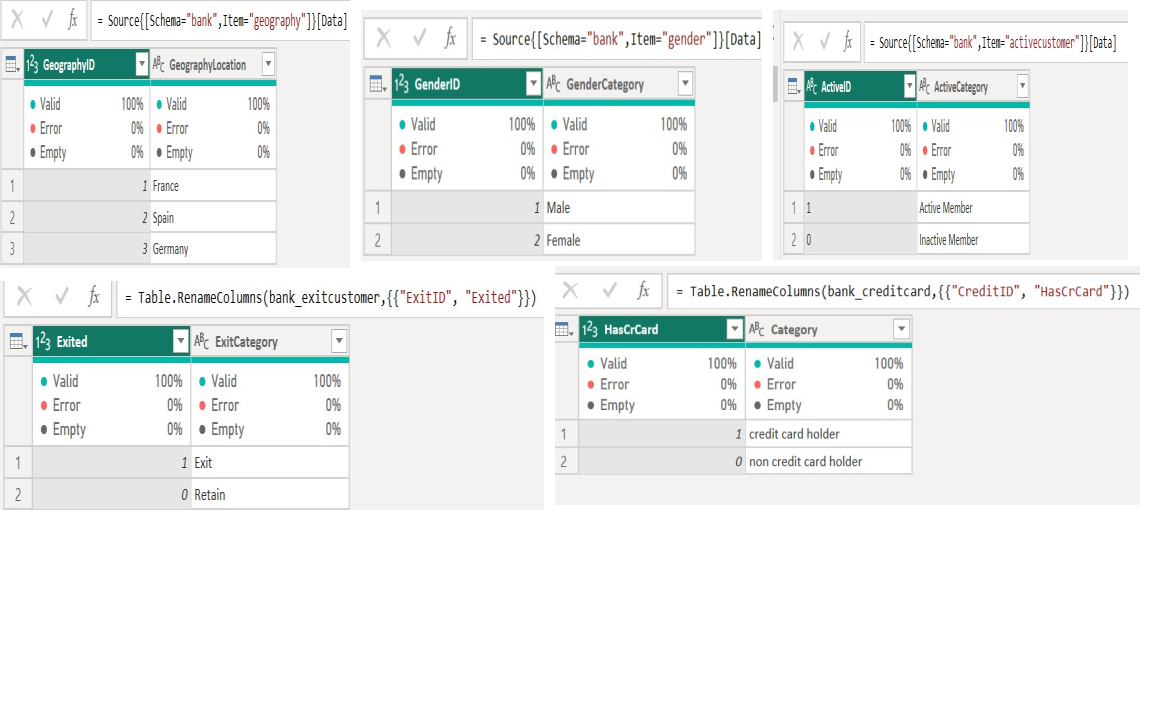


1. Were there any missing values in the data, using which tool did you replace them and what are the ways to handle them?



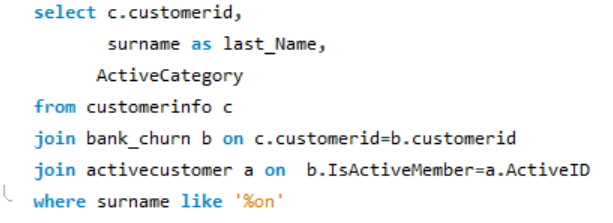
窗体底端





1. Write the query to get the customer ids, their last name and whether they are active or not for the customers whose surname  ends with “on”.

SQL query



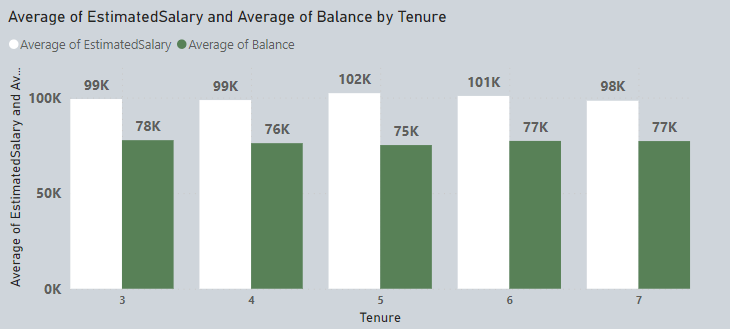
Output



**Subjective Questions**

1. **Customer Behavior Analysis:** What patterns can be observed in the spending habits of long-term customers compared to new customers, and what might these patterns suggest about customer loyalty?

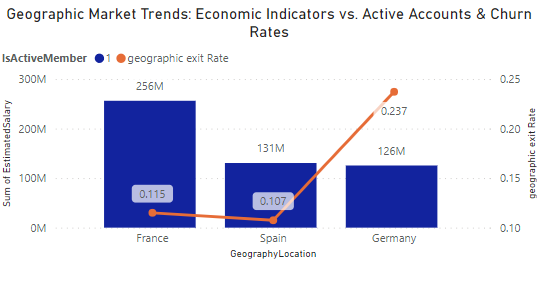
Examining the average estimated salary provides insight into customers' income diversity, indicating how much money they typically earn. Meanwhile, the average balance reflects the amount of money remaining in their accounts after expenditures, shedding light on their spending behaviors. Surprisingly, there aren't distinct disparities in spending habits between new and long-standing customers. Both groups exhibit similar patterns in their account balances, suggesting comparable spending behaviors.



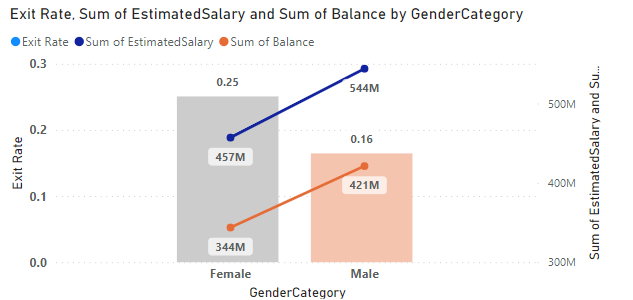
1. **Product Affinity Study:** Which bank products or services are most commonly used together, and how might this influence cross-selling strategies?

Without information on specific products or combinations, the data set lacks the necessary granularity to analyze relationships between products or propose cross-selling strategies. This limitation restricts the ability to uncover patterns such as which products are often purchased together or to identify opportunities for recommending complementary items to customers.

1. **Geographic Market Trends:** How do economic indicators in different geographic regions correlate with the number of active accounts and customer churn rates?

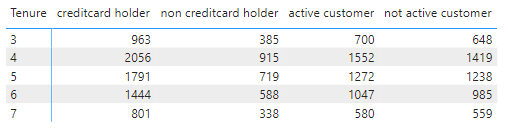
****The chart illustrates a clear correlation between estimated salary levels and account activity, as well as churn rates across different locations. In areas where estimated salaries are higher, there tends to be a greater number of active accounts and a lower churn rate. For instance, in France, where estimated salaries are high, there are more active customers and a lower churn rate. Conversely, in regions with lower estimated salaries like Germany, there are fewer active accounts and a higher churn rate, indicating more customers leaving the service. This pattern suggests that income levels play a significant role in customer engagement and retention within these locations.

1. **Risk Management Assessment:** Based on customer profiles, which demographic segments appear to pose the highest financial risk to the bank, and why?



The data reveals that a greater proportion of female customers are leaving the bank in comparison to male customers. Additionally, female customers generally exhibit lower estimated salaries and maintain lower account balances than their male counterparts. These indicators suggest that women may present a higher financial risk to the bank when compared to men.

1. **Customer Lifetime Value Forecast:** How would you use the available data to model and predict the lifetime (tenure) value of different customer segments?



From the visual representation and the data provided, it's evident that customers who possess a credit card play a pivotal role in generating interest income for the bank. This particular group of customers is expected to make a substantial and consistent contribution to the bank's revenue stream over time.

1. **Marketing Campaign Effectiveness:** How could you assess the impact of marketing campaigns on customer retention and acquisition within the dataset? What extra information would you need to solve this?

To evaluate the impact of marketing campaigns on customer retention and acquisition within the dataset, I would follow these steps:

* Segment customers: Split customers into distinct groups based on demographics, behavior, or other relevant criteria to analyze campaign effectiveness across different segments.
* Monitor campaign performance: Keep track of key campaign metrics like reach, engagement, and conversion rates over time to understand how well the campaigns are performing.
* Establish a control group: Form a control group comprising customers who weren't exposed to the campaign to serve as a benchmark, allowing for a comparison of the campaign's impact.
* Measure changes: Compare the behavior, retention, and acquisition rates of the exposed group with those of the control group to determine the campaign's effectiveness in driving these metrics.
* Analyze customer journey: Explore how customers interact with the campaign from initial awareness through to conversion, shedding light on its influence on retention and acquisition within each segment.
* Gather additional data: Collect supplementary information such as customer feedback and detailed campaign performance metrics to gain deeper insights into campaign effectiveness. Utilize historical data on past marketing campaigns and their outcomes to identify any discernible trends or patterns.

By following this approach, we can gain a comprehensive understanding of how marketing campaigns impact both customer retention and acquisition, enabling us to make informed decisions to optimize future marketing efforts.

1. **Customer Exit Reasons Exploration:** Can you identify common characteristics or trendsamong customers who have exited that could explain their reasons for leaving?

Analyzing customer churn solely based on limited data like credit scores and membership status can be challenging. While these metrics provide some understanding, they don't delve into the underlying reasons why customers leave.

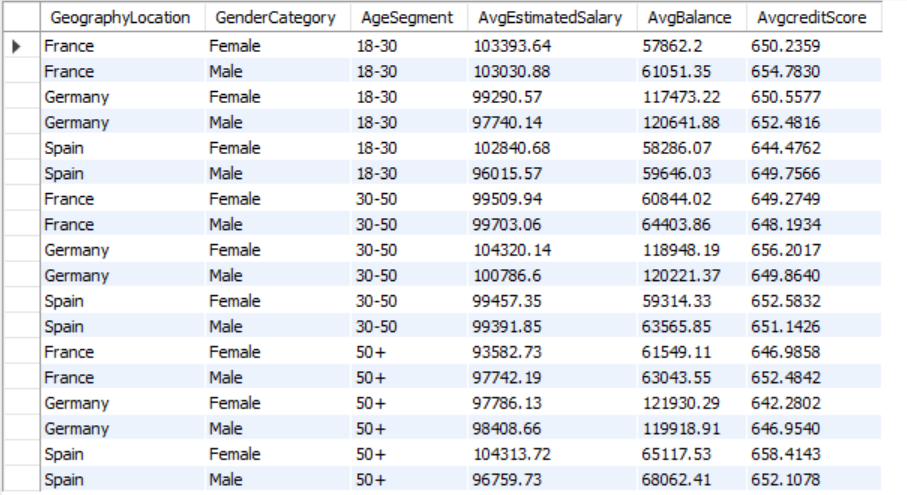
Customer attrition is influenced by various factors, including satisfaction l levels with products/services, experiences with customer service, competitive pressures, and external market dynamics.

To gain a comprehensive understanding and effectively manage churn, it's crucial to gather direct feedback from customers through surveys and meticulously examine their interactions. Augmenting existing data with sources such as customer feedback can provide invaluable insights into why customers are leaving and inform strategies to mitigate churn.

1. Are 'Tenure', 'NumOfProducts', 'IsActiveMember', and 'EstimatedSalary' important for predicting if a customer will leave the bank?

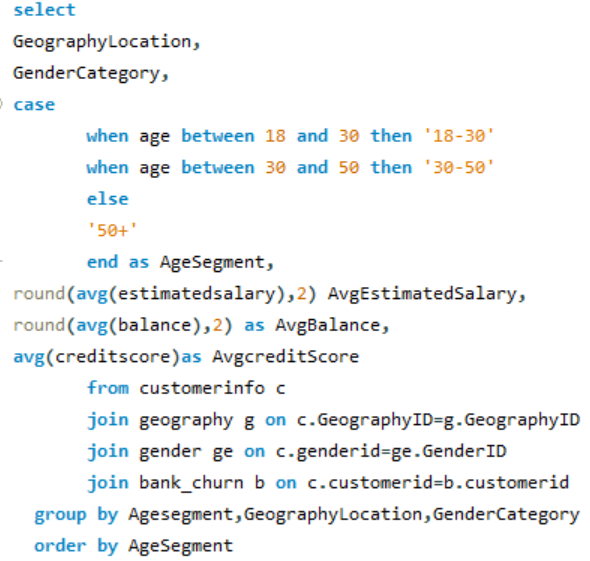
Predicting customer churn solely based on provided attributes poses challenges, as the dataset includes instances where even loyal and active customers have left the bank unexpectedly. Additionally, factors such as the number of products and estimated salary inconsistently explain customer attrition. To accurately forecast churn, it's essential to obtain direct feedback from departing customers. This feedback provides valuable insights into whether these attributes significantly contribute to predicting churn. Simply put, relying solely on numerical indicators like tenure or product count isn't sufficient to understand why customers leave. We must delve deeper into the reasons behind their decisions. Direct feedback becomes indispensable for refining churn prediction models and improving customer retention strategies.

1. Utilize SQL queries to segment customers based on demographics and account details.



Analyzing our customer data through demographic segmentation, I've examined age groups (18-30, 30-50, and 50+) and within each, scrutinized average estimated salary, account balance, and credit score across gender categories and locations. This detailed breakdown furnishes us with valuable insights into the financial behaviors and characteristics of our diverse customer segments.

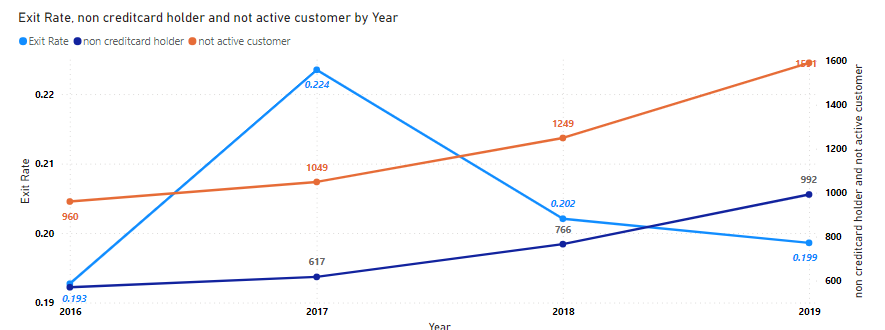
SQL query



1. How can we create a conditional formatting setup to visually highlight customers at risk of churn and to evaluate the impact of credit card rewards on customer retention?

* encountering obstacles in effectively visualizing customer churn data. While conditional formatting aids in identifying at-risk customers individually, it falls short in chart representation, hindering your ability to comprehend broader patterns such as the impact of demographics like location and gender on churn rates.
* Moreover, there's a notable gap in your analysis due to missing data on credit card rewards. This omission is significant as it's a crucial element in evaluating customer retention factors. Integrating this information is essential for gaining a comprehensive understanding of the underlying reasons behind customer churn. Addressing these challenges will be pivotal in enhancing the depth and accuracy of churn analysis insights.

1. What is the current churn rate per year and overall as well in the bank. Can you suggest some insights to the bank about which kind of customers are more likely to churn and what are the different strategies that can be used to decrease the churn rate.



Based on the analysis, it's evident that the churn rate peaked in 2017 but has since shown a slight decline. However, there's a concerning trend of increasing numbers of inactive customers and non-credit card holders each year, potentially contributing to rising churn rates. To address this, several strategies can be employed:

* Enhance Customer Engagement:

Tailor communication and marketing efforts to reconnect with inactive customers.

Provide proactive support to swiftly resolve any issues or concerns.

* Encourage Credit Card Adoption:

Promote the benefits of credit card usage to non-cardholders.

Offer incentives like discounts or promotions to encourage card usage and boost retention.

* Establish Feedback Channels:

Gather customer feedback through surveys to identify areas for improvement.

Utilize feedback to guide decision-making and prioritize initiatives aligned with customer preferences.

* Develop Loyalty Programs:

Introduce loyalty schemes to reward ongoing patronage and foster loyalty.

Provide exclusive perks or discounts to incentivize long-term relationships.

By implementing these strategies, the bank can mitigate churn rates, enhance customer satisfaction, and nurture lasting customer relationships. Continuous monitoring and adjustment of these initiatives are crucial for sustained success in reducing churn and bolstering customer retention.

1. How would you approach this problem, if the objective and subjective questions weren't given?

I'd kick off by delving into exploratory data analysis (EDA) to grasp the dataset's structure, variables, and distributions. My focus would be on key metrics like churn rate, customer demographics, active versus non-active accounts, and financial indicators. Through SQL queries, Excel, or Power BI, I'd dissect trends, correlations, and patterns within the data. SQL would come in handy for calculating churn rates by segment and examining customer demographics, while Excel's pivot tables and charts would visualize trends and relationships. With Power BI, I'd construct interactive dashboards to visually navigate the data and uncover actionable insights regarding customer behavior, product affinity, geographic trends, risk factors, and marketing effectiveness.

1. In the “Bank\_Churn” table how can you modify the name of “HasCrCard” column to “Has\_creditcard”?

In Power BI, we can easily rename a column by right-clicking on its name and selecting the option to rename it.

In SQL

ALTER TABLE bank\_churn

RENAME COLUMN HasCrCard TO Has\_creditcard;

Or

select hascrcard as Has\_creditcard from (

select \* from bank\_churn) as res;